

Reliability of Wound Measurement Methods

Dahlia Musa, Frank Guido-Sanz, Mindi Anderson and Salam Daher, *Member, IEEE*

Abstract: Reliable and accurate measurement methods are necessary for clinical assessment of wounds. Repeated measure of a wound indicates whether its healing is progressing or deteriorating, and if alternate treatment must be initiated. Many wound measurement techniques lack accuracy and reliability. **Technology:** We developed a software prototype that calculates 3-dimensional (3D) wound measurements from 3D scans. We conducted a study to compare the software prototype to physical and 2-dimensional (2D) image measurement techniques commonly used by clinicians. We compared inter-rater reliability between the techniques and measurements (i.e., length, width, depth, perimeter, surface area). **Results:** Inter-rater reliability was good or excellent for the physical, image, and software measurement techniques, however, there were significant differences in measurements between the techniques. For complex measurements (i.e., perimeter, surface area), the reliability of the software exceeded that of the physical and image techniques. **Conclusions:** Although inter-rater reliability was high for all measurement techniques, there was significant variability between the techniques. The software was overall most reliable, especially for calculation of complex measurements. **Clinical Impact:** Reducing the variability of wound measurements may improve patient outcomes, reduce the injury's prevalence, and mitigate the associated morbidity, mortality, and costs of these occurrences.

Index Terms—Wound measurement, 3-dimensional, 3D scanning, reliability, variability, physical, image, software

I. INTRODUCTION

WOUNDS require close monitoring to ensure that treatment is promoting adequate healing. The effectiveness of wound treatment is often determined by measuring the dimensions of the wound and assessing whether there is a reduction in its size over time. A 10-15 % reduction in wound size per week suggests increased likelihood of wound closure [1]. When a wound does not respond to a treatment, a new course of treatment may be initiated. Chronic wounds deviate from the typical progression of healing and are especially challenging to treat. These non-healing wounds are associated with increased morbidity and mortality, and present a financial burden to the healthcare system [2].

Pressure injuries are a common type of chronic wound [3]. Yearly, over 2.5 million Americans are affected by pressure injuries [4] and close to \$11 billion is spent in the United States on prevention and treatment for these injuries [5]. Despite the magnitude of this healthcare problem, preventing pressure injuries and implementing mitigating efforts remain difficult tasks. The Institute for Healthcare Improvement (IHI) made pressure ulcer prevention a top priority to support through the 5 Million Lives Campaign, aiming to reduce morbidity and mortality by implementing reliable science-based prevention measures [6]. More than ten years after the campaign, and despite being considered a quality-of-care indicator, pressure

injury incidence, prevalence, and elevated associated morbidity and mortality remain a top healthcare concern [7].

As important as identifying the wound history and location of the injuries during a physical exam, the physical characteristics of the pressure injuries (e.g., length, width, depth) need to be accurately assessed and documented [7]. Several barriers exist in current efforts to reduce the incidence and prevalence of pressure injuries, including the absence of a universal classification system to guide the management of injuries by stages [7] and the variability in methods to assess and measure the area of the injury. Several methods or techniques are used to size the injuries: ruler measuring, wound tracing with graduated acetate paper, and digital planimetry [8]. With these and other methods, an overestimation of wound size can occur [8], and the orientation or angle of the measurement can affect the accuracy of the sizing [9].

A common wound measurement technique used among clinicians is the ruler method. In the ruler method, the longest length (L) and widest width (W) are measured with a ruler, and the two measurements are multiplied as $L \times W$ to calculate the surface area [9]. The depth (D) is often measured by inserting a Q-tip into the wound and measuring the inserted segment of the Q-tip with a ruler [9]. The ruler method has gained prevalence because it is simple and quick, but it has disadvantages related to accuracy [9]. The $L \times W$ surface area equation measures the area of a rectangle, and since wounds are usually irregularly shaped, this equation results in a significant overestimation of about 41 % [10]. Some clinicians alternatively calculate the surface area as the area of an ellipse [11], but this measurement also assumes that the shape of the wound is elliptical. Additionally, the direct contact made with the wound is a possible source of infection and discomfort for the patient [12]. Several 2-dimensional (2D) digital measurement techniques have been developed to increase the accuracy of measurements and eliminate physical contact with the wound. In one method, ImageJ [13] software is used to calculate linear measurements between specified points on an image of the wound. While image measurements may be more accurate than ruler measurements, there are

This paper was submitted for review on 29 September 2022. Acknowledgment: This project has been partially funded by a grant from the New Jersey Health Foundation, and partially by the New Jersey Institute of Technology through PhD student Dahlia Musa. We thank Avkin for loaning us the simulated wounds measured in this study. We, the authors, report no conflicts of interest and are not affiliated with Avkin.

Dahlia Musa is a PhD student at the New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: dm79@njit.edu).

Dr. Frank Guido-Sanz is an assistant professor in the College of Nursing at the University of Central Florida, Orlando, FL 32816, USA (e-mail: Frank.Guido-Sanz@ucf.edu).

Dr. Mindi Anderson is a professor in the College of Nursing at the University of Central Florida, Orlando FL 328016, USA (e-mail: Mindi.Anderson@ucf.edu).

Dr. Salam Daher is an assistant professor in Informatics and Computer Science at the New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: salam.daher@njit.edu).

limitations. The angle of the camera may warp the image and effect the measurements. A 2D image also does not account for curvature, which can be problematic if the wound is located on a curved body part [14], such as the forearm or heel of the foot.

One alternative approach is to use 3-dimensional (3D) measurement systems. In these systems, measurements are taken on virtual 3D models that replicate the physical wound, and no direct contact is made with the wound. 3D models retain the surface geometry and color of the wound, and allow for manipulation, visualization, and enhanced clarity of wound features that is difficult to achieve in 2D images [15]. In many systems, 3D models are generated using structure from motion (SfM) photogrammetry [16, 17, 14, 18] or 3D scanners [19]. SfM matches features between 2D images and combines the data to create a 3D model. 3D scanners generate 3D models automatically without the need for 2D images. 3D scanners typically produce better quality 3D models than SfM; however, 3D scanners can be expensive and SfM is an affordable alternative [20]. 3D surface area measurements taken from 3D models have been shown to be more sensitive and thus better indicators of healing compared to 2D measurements [21]. Wound perimeter has also been recognized as an important factor in the prediction of wound healing [22]. 3D software can facilitate the measurement of perimeter as it is often difficult to calculate the perimeter of irregularly shaped wounds.

We developed a software prototype that calculates 3D wound measurements from 3D scans. The software can potentially improve clinicians' wound monitoring, assessment, and documentation in a non-contact method. This paper discusses the development and study of our software prototype. We address the following research questions:

- Q1: Does our 3D software support greater inter-rater reliability compared to physical and image measurement techniques?
- Q2: Does our 3D software support greater inter-rater reliability of complex measurements (i.e., perimeter, surface area) than simple measurements (i.e., length, width, depth)?
- Q3: Does variability exist between measurements (i.e., length, width, depth, perimeter, surface area) across different measurement techniques (i.e., physical, image, software)?

II. METHODS AND PROCEDURES

A. Development

1) 3D Wound Software Prototype

We developed the 3D wound software prototype in Unity 3D [23] using the programming language C#. The software takes 3D scans of the wounds as input and measures the length, width, depth, perimeter, and surface area of the wounds in a semi-automatic process. The user indicates the points of measurement by clicking on the 3D model, and the software calculates and displays the measurements between those points. Linear measurements are calculated by applying the Euclidean distance equation in 3D [24, 25]. The length, width, and depth are calculated as the distance between the end points indicated by the user. A linear measurement is shown in Figure 1A. Since the perimeter consists of several points around the

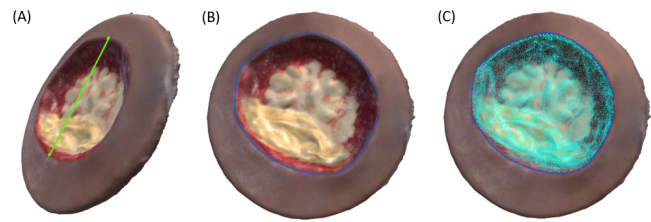


Fig. 1. (A) Linear measurement of a 3D wound model. (B) Perimeter measurement of a 3D wound model. (C) Surface area of a 3D wound model calculated automatically using the perimeter as the wound boundary.

wound edge, the perimeter measurement is calculated as the sum of the Euclidean distances between every two subsequent points. A perimeter measurement is shown in Figure 1B. For the surface area calculation, the perimeter points are used to define the boundary of the wound, and within this boundary, the software computes the sum of the mesh triangle areas. The surface area is calculated automatically from the perimeter points, but the user has the option to refine the measurement by adding or removing area from the calculation. A surface area measurement is shown in Figure 1C.

2) System Calibration

Equipment used in the calibration of the 3D wound software was metrologically traceable and calibrated. Calibration of the software was performed using the Scantech iReal 2E 3D scanner [26], Maya [27], and Unity [23]. The 3D scanner was calibrated according to the manufacturer's guidelines. The room temperature during calibration of the 3D scanner was within the range specified by the manufacturer (5°C to 40°C) [26]. After calibration of the 3D scanner, four Avkin [28] simulated wounds were 3D scanned individually with rulers and reference objects of known measurements placed in the x, y, and z directions. The rulers used in the 3D scans were compared with gauge blocks to ensure correctness of measurements. The 3D scans were imported into Maya, and using Maya's measurement tools, the 3D scans were scaled in the x, y, and z dimensions until the measurements of the reference objects in Maya were consistent with the objects' known measurements. The scale factor in Maya was found to be 0.1 in all dimensions. The mesh of the 3D models was then trimmed to remove the reference objects and include only the simulated wounds. The 3D models were exported from Maya as OBJ files and imported into Unity to be incorporated into the software. Maya and Unity both calculate measurements by applying the distance formula to two specified vertices, so calibration of the 3D models in Maya confirms calibration in Unity [25, 29]. We measured the diameter of the 3D models in our software to verify that the measurements were equivalent to the known diameter of the simulated wounds. This calibration process was performed separately for each of the four simulated wounds used in this study.

B. Participants

Participation was open to all students, faculty, and staff at New Jersey Institute of Technology (NJIT). Participants were excluded if they were not fluent in English or if they had vision impairment not corrected by glasses or contact lenses. Participants were not required to have any prior knowledge

of healthcare or wound measurement, but were expected to have basic knowledge of measurement and math. A total of 20 participants were included in the study. The data of two participants were excluded from analysis due to incompleteness, resulting in the inclusion of 18 participants' data. Participants' ages ranged from 18 to 31 years. Of the 18 participants, one participant had previous experience in wound measurement and all others had no previous experience.

C. Experiment Setup

The pilot study was conducted on the campus of NJIT. The physical and image measurements were taken using the NEIKO 01407A electronic digital caliper [30] with units set to millimeters (mm). The digital caliper was calibrated with gauge blocks to confirm correctness of measurements. The resolution of the caliper was 0.01 mm. The resolution of the wound images was 250 x 250 pixels. Based on the pixel dimensions of the wound images, the resolution of the image measurements was 0.48 mm. The mesh resolution of the 3D scans was 0.5 mm. The mesh resolution is the distance between the vertices of the mesh and determines the resolution of the software measurements. For the physical measurements, we placed four Avkin simulated wounds on a table, as shown in Figure 2A, and provided the caliper and Q-tips to measure the depth of the wounds. A webcam was positioned above the table to record participants as they measured the physical wounds. For the image measurements, we used color images of the wounds calibrated to account for camera distortions, and scaled them to their actual size. The images used for length, width, perimeter, and surface area were taken from a top-down view, and for depth, we took three images of each wound showing the inner edge of the wound from different angles. The images were printed onto standard white paper and placed on a desk for participants to measure with the caliper. A webcam was positioned on a tripod facing the desk to record participants as they measured the wound images. The setup for the image measurements is shown in Figure 2B. For the software measurements, the setup consisted of a desktop computer, mouse, and speakers. The software measurements were recorded in centimeters. The software setup is shown in Figure 2B. The wounds were positioned in the same orientation and presented in the same order for all measurement techniques.

D. Experiment Procedure

The procedure was approved by the NJIT Institutional Review Board. The study had a within-participant design. Before conducting the study, we received training from an expert clinician on the wound measurement techniques. Our instructions to participants were informed by our training. Participants were asked to measure the length, width, depth, perimeter, and surface area of four simulated wounds using the physical, image, and software measurement techniques. Participants used the caliper to take the physical and image measurements. Since the caliper could not be used to directly measure perimeter and surface area, these measurements required some calculation. We instructed participants to use the calculation method they believed was most suitable for the measurement estimation (e.g., approximating the wound shape to a simpler

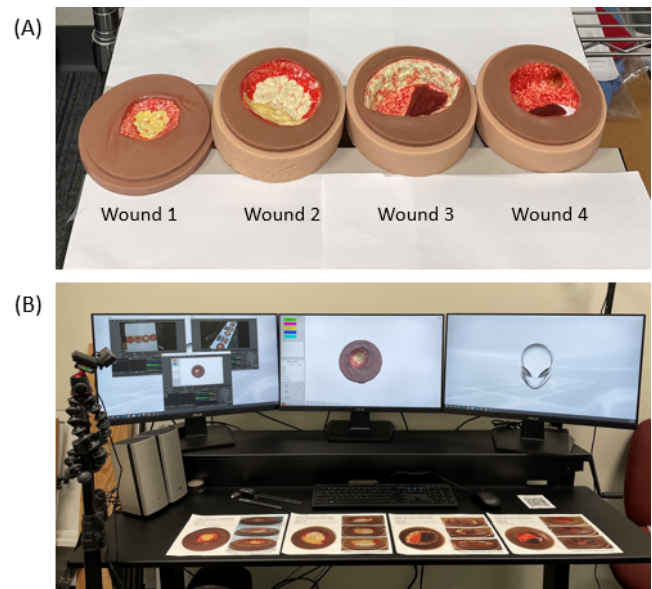


Fig. 2. (A) Setup of the Avkin simulated wounds for the physical measurement technique. (B) Setup of the wound images for the image measurement technique, and setup of the desktop computer for the software technique. The left monitor shows the view of the webcam recordings for each technique. The software interface is displayed on the middle monitor.

geometric shape and applying equations). Before taking the wound measurements, participants confirmed that they understood the measurement technique. As participants measured the wounds, we watched to ensure that they were using the instrument correctly. If we noticed participants incorrectly using the instrument, we corrected them and ensured that their measurements were ultimately taken through correct use of the instrument. Participants' demographic and measurement data were collected through Google Forms [31]. Participants accessed the online form on their cellphones using a QR code that we provided. Each participant submitted one measurement for the length, width, depth, perimeter, and surface area of each wound in each measurement technique, but they could take as many measurements as needed before submitting their final measurements. For all measurement techniques, participants were instructed to measure the length vertically and the width horizontally.

1) Physical Measurements

Participants watched as we showed them a live demonstration of how to use the caliper for measurements and the Q-tip technique to measure the wound depth. Participants then proceeded to measure the length, width, depth, perimeter, and surface area of the physical wounds. Participants entered their measurements into the online form.

2) Image Measurements

Participants watched as we showed them a live demonstration of how to measure the wound images with the caliper. For the depth measurement, participants were given three images depicting the wound depth from different angles. We instructed participants to choose the image they perceived to best represent the depth of the wound and measure the wound depth from that image. Participants then proceeded to measure the length, width, depth, perimeter, and surface area of the

wound images. Participants entered their measurements into the online form.

3) Software Measurements

Participants watched a short pre-recorded video that demonstrated how to use the software to measure the wounds. The first wound presented in the software was a practice wound to allow participants to gain familiarity with the software before measuring the four Avkin wounds. The practice wound was a physical simulated wound that we 3D scanned and imported into the software as we did with the Avkin wounds. Participants measured the length, width, depth, perimeter, and surface area of the practice wound and four Avkin wound 3D scans. Participants entered their measurements into the online form as they appeared in the software.

E. Statistical Analysis

During the study, we asked participants to measure the length vertically and width horizontally. However, many participants measured the length and width in the reverse or diagonally. To reduce inconsistency within the data, we defined the length as the greater measurement and width as the lesser measurement. We classified the numerical measurements as length and width according to this definition. Participants' original length and width measurements for the software technique are shown in Figure 3A, and the reclassified measurements are shown in Figure 3B. Before statistical analysis was performed, we also converted the software measurements from centimeters to millimeters since the physical and image measurements were recorded in millimeters. From this data, we analyzed the inter-rater reliability of the measurement techniques and variability between the techniques. Inter-rater reliability is used to measure agreement between raters [32]. High agreement between raters indicates that the measurement technique produces consistent results and is thus reliable. We analyzed inter-rater reliability as the absolute agreement, two-way mixed intra-class correlation coefficient (ICC) [33]. We analyzed variability between measurement techniques by comparing values of length, width, depth, perimeter, and surface area for each wound using the Friedman test and subsequent Conover post hoc tests. The Friedman test is a non-parametric test equivalent to the repeated measures ANOVA test [34]. We used a non-parametric test because the data were not normally distributed and therefore a parametric test is not recommended [35, 36]. Standard error was calculated as $\frac{\sigma}{\sqrt{n}}$, where σ represents the standard deviation and n represents the number of samples ($n = 18$).

F. Results

Descriptive statistics of the measurement data for the physical, image, and software measurement techniques are shown in Table I. Standard deviation was used to evaluate precision of the techniques. The lowest standard deviation in length measurements of wounds 2, 3, and 4 was observed from the software technique, and of wound 1 from the image technique. The lowest standard deviation in width measurements of wounds 1, 2, and 4 was observed from the software technique, and of wound 3 from the image technique. The lowest standard deviation in depth measurements of wound 2 was observed from the software technique, and of wounds 1, 3, and 4

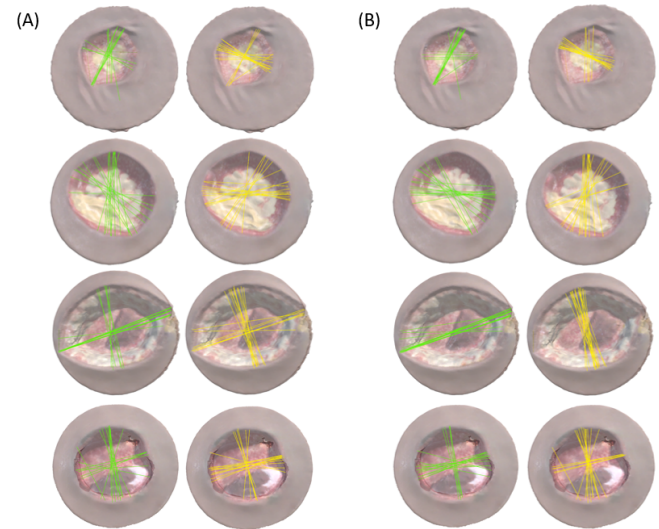


Fig. 3. (A) Participants' original length and width measurements recorded for the software technique. (B) Participants' length and width measurements for the software technique after reclassification. Length measurements are shown in green and width measurements are shown in yellow.

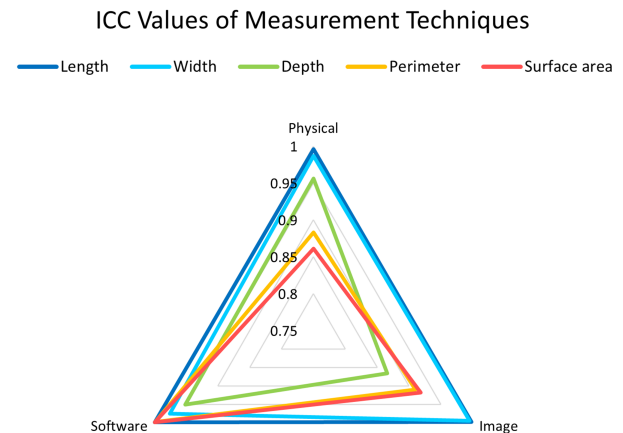


Fig. 4. Inter-rater reliability of measurement techniques analyzed as ICC values. These data are shown in Table II.

from the image technique. The lowest standard deviation in perimeter and surface area measurements of all wounds was observed from the software technique. Standard error values followed the same trend reported for standard deviation.

Inter-rater reliability. ICC values were greater than 0.8 for length, width, depth, perimeter, and surface area measurements for the physical, image, and software measurement techniques, indicating high reliability of all three techniques for all measurements. ICC values greater than 0.9 were observed for physical length, width, and depth measurements; for image length, width, perimeter, and surface area measurements; and for all software measurements. The highest ICC value observed was 0.99 for software length and surface area measurements. The lowest ICC values observed were 0.861, 0.866 and 0.883 for physical surface area, image depth, and physical perimeter, respectively. The inter-rater reliability results are shown in Figure 4 and Table II.

Length variability. There were statistically significant ($p < 0.05$) differences in length measurements between the

	Wound	Physical		Image		Software		Friedman	Conover post hoc		
		Mean \pm SE	SD	Mean \pm SE	SD	Mean \pm SE	SD		P vs. I	P vs. S	I vs. S
Length	1	51.10 \pm 1.11	4.73	52.39 \pm 0.74	3.16	53.49 \pm 0.88	3.75	0.042	0.143	0.017	0.324
	2	71.52 \pm 1.02	4.31	77.70 \pm 0.54	2.29	73.83 \pm 0.36	1.54	< .001	< .001	0.251	< .001
	3	94.36 \pm 1.49	6.30	100.60 \pm 1.20	5.08	97.29 \pm 1.02	4.35	< .001	< .001	0.054	0.054
	4	60.89 \pm 0.84	3.57	68.55 \pm 0.74	3.14	60.57 \pm 0.49	2.07	< .001	< .001	0.741	< .001
Width	1	46.35 \pm 0.92	3.91	49.23 \pm 0.84	3.55	48.67 \pm 0.79	3.35	0.002	0.002	0.020	0.362
	2	67.48 \pm 0.96	4.05	74.45 \pm 0.90	3.82	69.72 \pm 0.31	1.30	< .001	< .001	0.105	0.003
	3	63.60 \pm 1.22	5.18	69.88 \pm 0.73	3.09	64.55 \pm 2.68	11.39	< .001	< .001	0.012	0.012
	4	57.39 \pm 1.13	4.78	65.24 \pm 0.75	3.18	58.17 \pm 0.72	3.04	< .001	< .001	0.620	< .001
Depth	1	13.54 \pm 1.00	4.26	13.94 \pm 0.96	4.09	13.73 \pm 1.18	4.99	0.211	0.191	0.741	0.105
	2	20.13 \pm 1.77	7.50	20.12 \pm 2.05	8.71	17.90 \pm 1.96	8.31	0.115	0.620	0.054	0.143
	3	30.06 \pm 1.54	6.52	20.96 \pm 0.74	3.15	32.09 \pm 2.48	10.54	< .001	< .001	0.410	< .001
	4	25.82 \pm 1.66	7.05	20.36 \pm 0.72	3.04	21.09 \pm 0.98	4.18	0.006	0.012	0.008	0.869
Perimeter	1	152.95 \pm 13.61	57.76	162.18 \pm 8.61	36.54	160.25 \pm 0.75	3.18	0.078	0.036	0.120	0.561
	2	203.11 \pm 11.26	47.77	226.73 \pm 12.13	51.46	229.48 \pm 2.51	10.67	< .001	< .001	0.017	0.143
	3	238.35 \pm 15.30	64.90	258.03 \pm 15.73	66.74	257.81 \pm 2.37	10.06	0.042	0.017	0.143	0.324
	4	188.81 \pm 9.32	39.54	208.77 \pm 12.43	52.72	198.76 \pm 1.52	6.47	0.003	0.002	0.251	0.037
Surface area	1	2389.23 \pm 532.35	2258.59	2102.95 \pm 153.99	653.34	2163.67 \pm 25.90	109.88	0.066	0.076	0.037	0.741
	2	3593.45 \pm 262.23	1112.55	4493.91 \pm 471.90	2002.10	4892.76 \pm 24.29	103.06	< .001	0.017	< .001	0.143
	3	5950.36 \pm 905.50	3841.71	5904.89 \pm 635.13	2694.64	7986.66 \pm 52.67	223.44	< .001	0.191	< .001	0.008
	4	3349.20 \pm 413.07	1752.49	4250.97 \pm 452.97	1921.78	6219.16 \pm 108.94	462.17	< .001	0.017	< .001	0.030

TABLE I

RESULTS OF THE FRIEDMAN AND CONOVER TESTS EVALUATING VARIABILITY BETWEEN PHYSICAL (P), IMAGE (I), AND SOFTWARE (S) MEASUREMENT TECHNIQUES. DESCRIPTIVE DATA INCLUDE MEAN, STANDARD ERROR (SE), AND STANDARD DEVIATION (SD). UNITS FOR LENGTH, WIDTH, DEPTH, AND PERIMETER ARE IN MM, AND UNITS FOR SURFACE AREA ARE IN MM². STATISTICALLY SIGNIFICANT P-VALUES ($P \leq 0.05$) ARE IN BOLD TEXT. THESE RESULTS ARE GRAPHED IN FIGURE 5.

Measurement	ICC values		
	Physical	Image	Software
Length	0.996	0.998	0.999
Width	0.987	0.995	0.975
Depth	0.956	0.866	0.951
Perimeter	0.883	0.910	0.998
Surface area	0.861	0.918	0.999

TABLE II

INTER-RATER RELIABILITY OF MEASUREMENT TECHNIQUES ANALYZED AS ICC VALUES. ALL COMPARISONS HAVE A P-VALUE OF < .001. THESE RESULTS ARE DEPICTED AS A KIVIAIAT DIAGRAM IN FIGURE 4.

physical, image, and software measurement techniques for wounds 1, 2, 3, and 4. Pair-wise comparisons revealed statistically significant differences between the physical and image techniques for wounds 2, 3, and 4; between the physical and software techniques for wound 1; and between the image and software techniques for wounds 2 and 4. These results for the Friedman and Conover tests are shown in 5A and Table I.

Width variability. There were statistically significant differences in width measurements between the physical, image, and software measurement techniques for wounds 1, 2, 3, and 4. Pair-wise comparisons revealed statistically significant differences between the physical and image techniques for wounds 1, 2, 3, and 4; between the physical and software techniques for wounds 2, 3, and 4; and between the image and software techniques for wounds 2 and 4. These results for the Friedman and Conover tests are shown in 5B and Table I.

Depth variability. There were statistically significant differences in depth measurements between the physical, image, and software measurement techniques for wounds 3 and 4. Pair-wise comparisons revealed statistically significant differences between the physical and image techniques for wounds 3 and 4; between the physical and software techniques for wound 4; and between the image and software techniques for wound 3.

These results for the Friedman and Conover tests are shown in 5C and Table I.

Perimeter variability. There were statistically significant differences in perimeter measurements between the physical, image, and software measurement techniques for wounds 2, 3, and 4. Pair-wise comparisons revealed statistically significant differences between the physical and image techniques for wounds 1, 2, 3, and 4; between the physical and software techniques for wound 2; and between the image and software techniques for wound 4. These results for the Friedman and Conover tests are shown in 5D and Table I.

Surface area variability. There were statistically significant differences in surface area measurements between the physical, image, and software measurement techniques for wounds 2, 3, and 4. Pair-wise comparisons revealed statistically significant differences between the physical and image techniques for wounds 2 and 4; between the physical and software techniques for wound 1, 2, 3, and 4; and between the image and software techniques for wounds 3 and 4. These results for the Friedman and Conover tests are shown in 5E and Table I.

G. Discussion

During our study, many participants misrepresented the length and width of the wounds despite instructions to measure the length vertically and width horizontally. Participants' inconsistent classifications of length and width suggest that measurement error may result from misunderstanding or lack of clear instruction. It is likely that clinical wound measurements are susceptible to the same errors; therefore, it is important that visual representations of clinicians' measurements are documented in patients' records and that measurements can be repeated if errors are found. Since physical measurements are performed directly on the patient, measurements cannot be repeated at a later point in time under the same conditions.

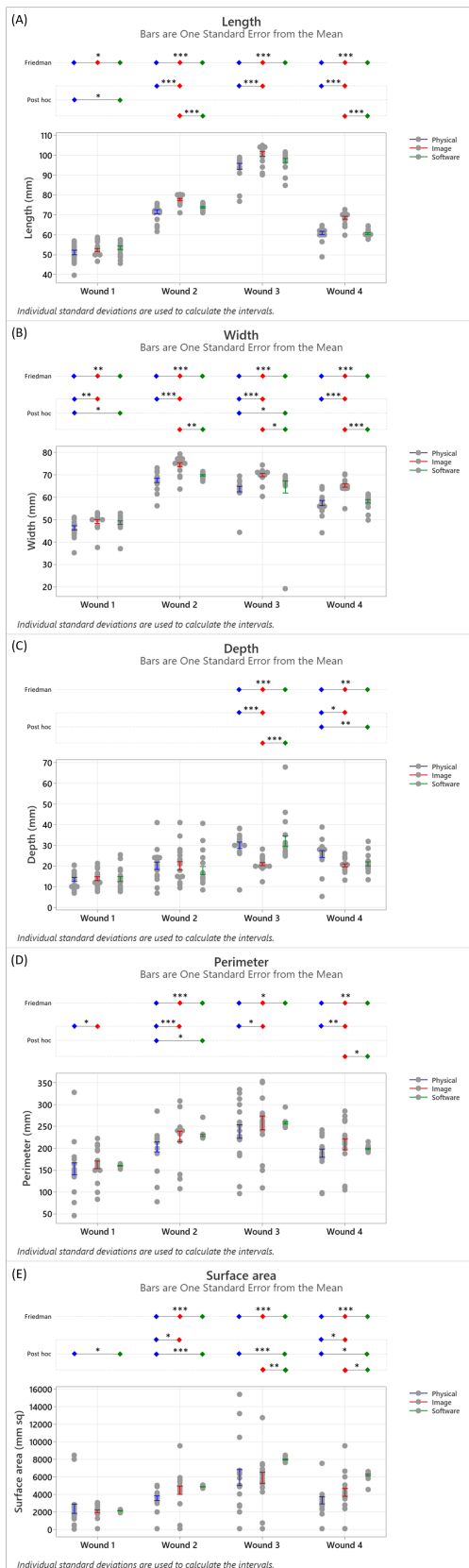


Fig. 5. Individual value plots showing the distribution of participants' physical, image, and software measurements of each wound for (A) length, (B) width, (C) depth, (D) perimeter, and (E) surface area. Error bars represent one standard error from the mean. These data are shown in Table I. (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$)

Clinicians may retain the original image used for image measurements, but images are inherently prone to distortion and warping and do not accurately depict curved surfaces [37]. Our 3D software maintains 3D models that accurately represent wound surface geometry and color, as well as visual indicators of the measured distances. The software measurements can also be repeated on the same 3D model as needed. In addition to the advantages of visualization and repeatability of measurements, our results suggest that the software may have greater reliability and reduced variability compared to physical and image measurement techniques.

Inter-rater reliability of measurement techniques. ICC values between 0.75 and 0.9 indicate good inter-rater reliability, and ICC values above 0.9 indicate excellent reliability [33]. The ICC values for all measurement techniques suggested good and excellent reliability, however, the reliability of the software technique was consistently excellent for all measurements. Overall, the software supported greater inter-rater reliability compared to the physical and image measurement techniques. These results answer our first research question (Q1).

Inter-rater reliability of simple and complex measurements. Comparison of ICC values between measurements revealed a noticeable trend between simple measurements (i.e., length, width, depth) and complex measurements (i.e., perimeter, surface area). ICC values for simple measurements were excellent across all techniques except image depth, which yielded an ICC value of 0.866. The lower ICC value for image depth is unsurprising as depth information is difficult to represent in a 2D image. The reliability of the physical and image techniques was decreased for complex measurements compared to simple measurements, while the reliability of the software technique was increased for complex measurements compared to simple measurements. In the physical and image techniques, complex measurements cannot be measured directly and instead must be calculated through use of a geometric equation or other methods. Many participants reported in the survey that for the physical and image techniques, they calculated the perimeter and surface area by applying equations for circle circumference and area, respectively. The software, however, did not require any calculation as the complex measurements were computed automatically from participants' inputs. The software reduced variability in complex measurements compared to the physical and image techniques by automating the computation. These results suggest that the software is more reliable than the physical and image techniques for complex measurements. These results answer our second research question (Q2).

Variability between measurement techniques. When comparing measurement values between techniques, we found that there were significant differences, primarily between the physical and image techniques. In the physical technique, measurements were taken on the actual wounds in 3D space, while in the image technique, measurements were taken in flat 2D space. These results suggest that the loss of a dimension in 2D images affects the measurements. Despite the high inter-rater reliability of the physical and image measurement techniques, this variability suggests that the techniques are not equivalent. Inter-rater reliability for these techniques were

likely high because the techniques were easy to use, but the reliability is not indicative of the techniques' accuracy. In fact, the occurrence of both high inter-rater reliability and variability between the techniques suggests that the techniques may be inherently susceptible to error. Since the focus of this study was on reliability, we have no further analysis on the accuracy of the measurement techniques, but we plan to investigate accuracy in future studies. Nevertheless, the variability between measurement techniques is concerning, and this variability may also exist in clinical practice. In clinical practice, physical and image measurement techniques are frequently used to monitor wound healing. Depending on which technique is used, the patient's healing trajectory may be interpreted differently and influence treatment decisions. To ensure that patients receive the most effective wound care, it is essential that measurement techniques are consistent. These results answer our third research question (Q3).

Accurate and consistent measurement of wounds can facilitate tracking the progress or deterioration of healing, allowing clinicians to assess the injury and evaluate the effectiveness of the treatment [9]. Reducing the variability in measurements of wound size and characteristics (e.g., depth) may improve patient outcomes, reduce the injury's prevalence, and mitigate the associated morbidity, mortality, and costs to these occurrences. The proposed technology has the potential to benefit patients, their families, and health systems. Moreover, this proposed technology may facilitate the recording and exchange of data between members of the team caring for patients.

III. LIMITATIONS

Our study was limited by three factors. Firstly, the inconsistencies in participants' classifications of length and width forced us to define the measurements on different terms. In clinical practice, the length is defined as the greatest measurement along the patient's head-to-toe axis, and the width is the greatest measurement perpendicular to the length [9]. By defining the length as the longest vertical measurement and width as the widest horizontal measurement, our definitions would have been consistent with those used by clinicians. Our reclassification of length as the larger measurement and width as the smaller measurement were still valid definitions of length and width, though not practiced by clinicians. The second limitation was that we did not have ground truth values to compare participants' measurements to and as a result, we were unable to assess the accuracy of the measurement techniques. The simulated wounds used in this study were handmade without any pre-defined dimensions. Any ground truth values would have been measured using one of the techniques evaluated in this study, and therefore may have biased our results. Although data pertaining to measurement accuracy may have enhanced our results, the lack of this data did not affect the findings reported in this paper. Lastly, participants of this study were not healthcare professionals and did not have a background in wound measurement. This was a pilot study intended to test the capability of the 3D wound software. We could have asked participants to measure any objects, but we used simulated wounds in anticipation of future studies that will involve healthcare professionals. We

plan to compare these results to those we will obtain from healthcare professionals.

IV. CONCLUSION

In this paper, we proposed a safe and reliable software capable of measuring wounds in 3D. We compared our 3D software to physical and 2D image measurement techniques commonly used in clinical practice. Inter-rater reliability was high for all techniques; however, measurements between the techniques differed significantly, indicating that the techniques did not produce equivalent measurements. Additionally, the software was more reliable than the physical and image techniques for computation of complex measurements. Based on these results, we recommend that clinicians consider software as an alternative to physical and image measurement techniques to reduce variability of wound measurements and ultimately deliver more effective wound treatment to patients.

REFERENCES

- [1]
- [2]
- [3]
- [4]
- [5]
- [6]
- [7]
- [8]
- [9]
- [10]

- | | |
|------|------|
| [11] | [23] |
| | [24] |
| [12] | [25] |
| | [26] |
| [13] | |
| [14] | [27] |
| | [28] |
| [15] | [29] |
| | |
| [16] | [30] |
| | [31] |
| [17] | [32] |
| | |
| [18] | [33] |
| | |
| [19] | [34] |
| | [35] |
| [20] | [36] |
| | [37] |
| [21] | |
| | |
| [22] | |